Variables Discretas y Continuas

Thursday, January 31, 2019 9:29 AM

- Variables Continuas
  - Son infinitas
  - Por lo regular son magnitudes físicas.

Discretización

- Variables Discretas
  - Son finitas o infinitas
  - Consta de unidades de unidades o partes que no estan separadas unas de otras.

Aspectos generales

- La población es todo el conjunto a evaluar
- La muestra es una pequeña parte de la población (Subconjunto)
  - ↳ Son tomadas mediante métodos aleatorios

Estadística Descriptiva

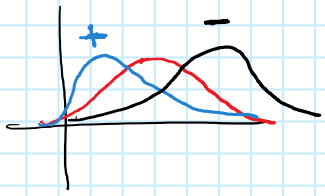
Thursday, January 31, 2019 9:49 AM

- Sirve para describir a través de medidas:

→ Tendencia Central
- Media
- Mediana
- Moda

→ Dispersión
- Varianza
- Desviación
  - estándar
  - media
- Coeficiente de variación

→ Forma
- Coeficiente de sesgo (Derecha o izquierda)
- Coeficiente de curtosis (Grado de picudez)

- 

//En dirección de la colita

Positivo → Izquierda

-

- Platicúrtica
  ↳ Varianza y desviación estandar
  ↳ muy grandes
- Leptocúrtica*
- Mesocúrtica

* Con base a la Normal "0" y "3"

- Se pueden usar:
  - Medidas descriptivas
  - Tablas de frecuencias
  - Gráficas

Nomenclatura

- Medidas de población → Letras griegas minúsculas
- Respecto a la muestra → Letras griegas (mayúsculas) → Variable
  → Letras griegas (minúsculas) → valor de la variable

Medidas

- Agrupados / Sin agrupar
  ↳ Determinar cantidades en rangos

- **Promedio** / **Media** : $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$

- **Mediana** :

$$\tilde{X} = \begin{cases} \dfrac{X\left(\frac{n}{2}\right) + X\left(\frac{n}{2}+1\right)}{2} & \leftarrow \text{Parentesis; Par} \\ X\left(\frac{n+1}{2}\right) & ; \text{Impar} \end{cases}$$

- **Moda** : Valor más repetido

- Varianza: $\sigma^2 = E[(x-\mu)^2] f(x)$

  si $f(x) = \frac{1}{N}$

  - Si es una población finita

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{n} (x_i - \mu)^2$$

- Desviación estandar: $\sigma = \sqrt{\sigma^2}$

- Desviación media: $= \frac{\sum |x_i - \mu|}{N}$

- Coeficiente de variación: $CV = \frac{s}{\bar{x}} \triangleq \frac{\text{Desviación estandar}}{\text{promedio}}$

- Fractiles o cuantiles
  - Cuartiles { 4 partes - 25%
  - Mediana { 2 partes - 50%
  - Deciles { 10 partes - 10%
  - Centiles o percentiles { 100 partes - 1%

* Dividimos al conjunto

TAREA: Relación entre "media mediana o moda"

TAREA: Relación entre "media, mediana y moda"

- Cuando una distribución de frecuencia es simétrica la media, mediana y moda coinciden en su valor

$$\bar{x} = M_e = M_o$$

- En una distribución sesgada a la izquierda:
  - la moda es menor a la mediana y esta a su vez menor que la media

$$\bar{x} < M_o < M_e$$

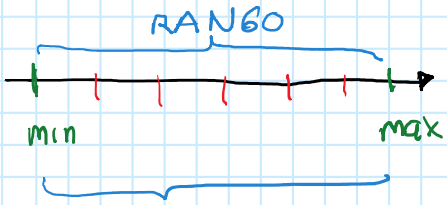
- En una distribución sesgada a la derecha
  - la moda es mayor a la mediana y esta a su vez mayor que la media

$$\bar{x} > M_o > M_e$$

Datos Agrupados

Thursday, February 7, 2019 9:14 AM

- Formas de describir datos:
  - Medias descriptivas
  - Tablas de Frecuencia
  - Gráficas



Las divisiones se les conoce como CLASES

- rango = max − min

- Longitud de Clases $\triangleq$ "c"
- Número de clases $\triangleq$ "m"

- Longitud Clase $\triangleq c = \dfrac{\text{rango}}{m} = *$

* Debemos redondearla a 3 o 3.5

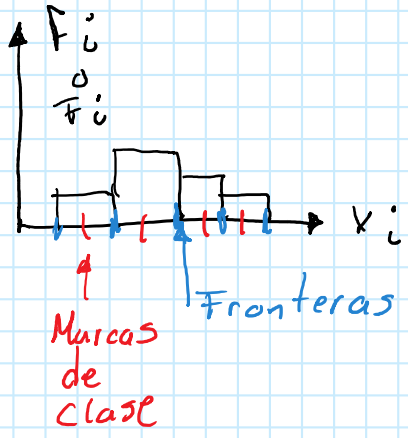
// Como se excede en un pequeño tramo debemos comparar el rango viejo con el nuevo y posteriormente centrarlos.

* Digitos significativos → 1 decimal más

- Intervalos de Clase | Limites Aparentes
- Fronteras de Clase | Limites Reales

Gráficas
- Histogramas
- Polígonos
  - Frecuencias (Marcas de Clase) $\{ X_i$ vs f_i o F_i
  - Frecuencias relativas o acumuladas $\{$ Fronteras vs F'_i

- Frecuencia $- f_i$; f'_i - relativa
- Frecuencia Acumulada $- F_i$; F'_i - relativa



Frecuencia de datos

- Reglas para la determinación de cantidad de clases:

- Regla de Sturges : Fue enunciado en 1926 por el matemático alemán Herbert Sturges

Sturges : matemático alemán
Herbert Sturges

$$K = 1 + \log_2(N)$$

- Regla de "raíz de n" : • Es una propuesta empírica donde n es la cantidad de datos

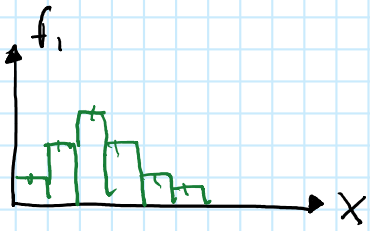
$$\sqrt{n}$$

Histogramas - Polígonos

Thursday, February 14, 2019 9:24 AM

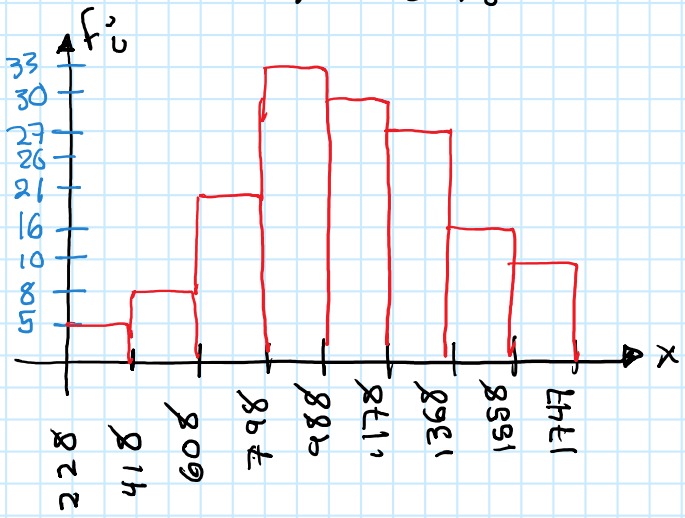
Histogramas

- Son gráficas de barras
- X: $V_X = \begin{cases} f_i \\ f'_i \\ F_i \\ F'_i \end{cases}$



//Barras juntas

Gráfica X vs f_i



Polígonos de Frecuencia

- X_i vs $\begin{cases} f_i \\ f'_i \end{cases}$
- Ojiva es un polígono de frecuencia acumulada
- Fronteras vs $\begin{cases} F_i \\ F'_i \end{cases}$

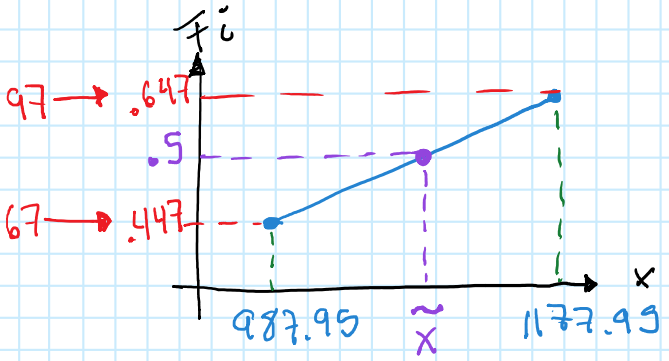
Interpolación

$$y_x = y_0 + \frac{x - x_0}{x_1 - x_0}(y_1 - y_0)$$

Fractiles

$$\text{Fractil} = L_{inf} + \left[\frac{\frac{n}{2} - F_{i-1}}{f_i}\right] c_i$$

- $n \triangleq$ número de datos
- $F_{i-1} \triangleq$ Frecuencia acumulada de la clase anterior



$$\text{Fractil} = L_{inf} + \left[\frac{\frac{n}{2} - F_{i-1}}{f_i}\right] c_i =$$

$$= 987.95 + \left[\frac{\frac{150}{2} - 67}{30}\right](190) =$$

$$= 1038.6166$$

▶ Que "$x < 1000$" y "$x > 1300$"

$$P(x < 1100 \cup x > 1300) = 1 - P(1100 < x < 1300) =$$

$$= P(x < 1100) + [1 - P(x < 1300)]$$

EXAMEN

Thursday, February 21, 2019 9:14 AM

* muestra

- cantidad = 100 automoviles
- Longitud Clase = 20 autos

	Marcas clase X_i	f_i	F_i	f'_i	F'_i
29.5 – 49.5	39.5	12	12	.12	.12
49.5 – 69.5	59.5	36	48	.36	.48
69.5 – 89.5	79.5	30	78	.3	.78
89.5 – 109.5	99.5	20	98	.2	.98
109.5 – 129.5	119.5	2	100	.02	1

- Moda $= 49.5 + \left[\dfrac{24}{24+6}\right](20) = 65.5$

$a = 36 - 12 = 24$

$b = 36 - 30 = 6$

- Varianza

$$= \frac{1}{n-1}\left[\sum_{i=1}^{m} f_i (x_i)^2 - n\,\bar{x}^2\right] =$$

$$= \frac{1}{99}\left[(12)(39.5)^2 - 100(72.3)^2\right]$$

$(36)(59.5)^2$

$(30)(79.5)^2$

$(20)(99.5)^2$

$$(20)(99.5)^2$$
$$(2)(119.5)^2$$

18723
127449
189607.5
198005
28560.5

$$= \frac{1}{99}\left[562345 - 100(72.3)^2\right] = 400.1616162$$

Desv. est = 20.00404

- Segundo Cuartil

$$\text{Cuartil} = 69.5 + \left[\frac{100\left(\frac{1}{2}\right) - 48}{30}\right](20) =$$

$$= 70.83$$

Estadística Inferencial

Friday, February 22, 2019 8:53 PM

- La estadística inferencial es la que nos proporciona los métodos y técnicas para hacer generalizaciones y estimaciones que nos permiten predecir mediante modelos matemáticos el comportamiento futuro de un hecho en estudio.

→ Con base a la estadística descriptiva.

Conceptos Básicos de Inferencia

- Población → Muestra

* La muestra debe ser tomada aleatoriamente
* La población son hipotéticas

* Población es el conjunto de las medidas a considerar.

Ejemplo: Tiempo de carga de las PC
- → Unidad Experimental → PC
- → Población (Muestra) → Tiempos asociados

Variables

Discretas
- Son aquellas que no pueden tomar cualquier valor
- //Números enteros asociados

Continuas
- Son por lo regular magnitudes Físicas
- Nunca puede ser medida con exactitud

Parametros

* Se denotan letras griegas minúsculas

NOTA | Muestreo $\triangleq$ Proceso de tomar una muestra de una población.

Fundamento de Estadística Inferencial
- Distribución de probabilidad de cada observación es la misma a la de la población.
- Siempre es la misma en toda la población.

Distribución de Probabilidad

- $L(x_1, x_2, x_3 \ldots x_n) = f(x_1) \cdot f(x_2) \cdot f(x_3)$
- Función de verosimilitud de la muestra

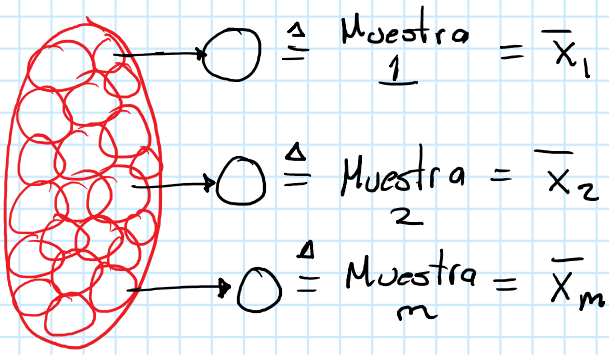
$$L(\underline{x}) = \prod_{i=1}^{n} f(x_i)$$

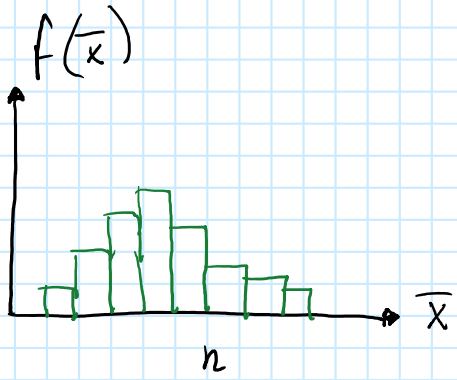
Estadísticos *

- Es una función de las observaciones de la muestra
  // No contiene cantidades desconocidas

Población
↓
Muestra
↓
Observación

- Total de la muestra $\triangleq T = \sum_{i=1}^{n} x_i$
- Media muestral $\triangleq \bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$





Distribución de muestreo de estadístico → • Distribución de Probabilidad del Estadístico $\bar{X}$

- Estadístico es una variable aleatoria
  ↳ Tiene una distribución
    ↳ Media, Varianza

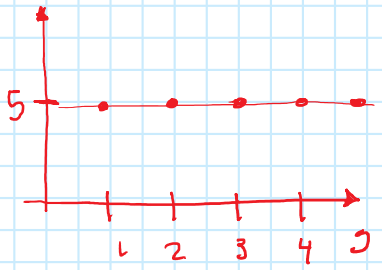
Estimador *

- Es un estadístico que se utiliza para inferir el valor de algún parámetro poblacional y que proporciona un único valor.

// Media muestral ← sirve para estimar la media de la población

---

* Fundamentos de Inferencia Estadística



• Distribución Uniforme Discreta

• La media de las medias de los estadísticos es igual a la media de la población

↳ $E[\bar{x}] = \mu$

• La varianza de la población es igual a la varianza de la muestra dividida entre su tamaño

// Tamaño de muestra grande → Menos variabilidad

↳ $V[\bar{x}] = \frac{\sigma^2}{n}$

▸ Propiedad Aditiva de la Distribución Normal

• Una combinación de variables aleatorias me da como resultado una Distribución Normal

• Con base a lo anterior "inferimos" que la forma obtenida también es Normal

$x_i \triangleq$ Forma Normal $\qquad \bar{x} = x_1 + \ldots + x_i$

∴ La media de la muestra tambien es Normal

1] Sea una población con distribución normal, con media 30 y varianza 16, si se toman muestras de tamaño n=16,

a) Determinar la distribución de muestreo de la media e indicar cuáles son sus parámetros (media y varianza).

1] Sea una población con distribución normal, con media 30 y varianza 16, si se toman muestras de tamaño n=16,

a) Determinar la distribución de muestreo de la media e indicar cuáles son sus parámetros (media y varianza).
b) encontrar la probabilidad de que al sacar una muestra la media de la muestra no difiera de la media poblacional en más de 2.0

a) $\mu_{\bar{x}} = \mu$ y $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

$\bar{x} \sim N(30, \frac{16}{10})$; $\bar{x} \sim N(30, 1.6)$ //

b) $P(|\bar{x} - \mu| < 2) = P(-2 < \bar{x} - \mu < 2) =$

$= P\left(-\frac{2}{\left(\frac{4}{\sqrt{10}}\right)} < \frac{x-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} < \frac{2}{\left(\frac{4}{\sqrt{10}}\right)}\right) \cong$

$\cong P(-1.58 < z < 1.58) = P(z < 1.58) - P(z < -1.58) =$

$= .9429 + .9429 = 1.8858$

$= 1.8858 - 1 = .8858$ //

Aditividad en Distribución Normal

- $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu, \sigma_Y)$

$\therefore X + Y \sim N(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$

NOTA: X y Y son independientes, entonces la distribución de la suma o diferencia de ambas variables es también Normal

Ejercicios

Tuesday, March 5, 2019 9:32 AM

Distribuciones de Probabilidad

Distribución de muestreo de la media muestral

$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

- Población Normal (Distribución Normal) $\rightarrow \bar{x} \sim N(\mu, \frac{\sigma^2}{n})$
- Otra distribución
  - Si $n < 30$; $\bar{x} \sim$ Desconocida
  - Si $n > 30$; Por el teorema del limite central $\bar{x} \dot{\sim} \text{Normal}(\mu, \frac{\sigma^2}{n})$

PROBABILIDAD

- Bernoulli $\triangleq$ Regreso Objeto
- Hipergeométrica $\triangleq$ No regreso

- $\sigma^2_x = \frac{\sigma^2}{n} \triangleq$ Población grande

---

2]

a) $E[\bar{x}] = \mu$, $E[\bar{x}] = 500$

b) $V[\bar{x}] = \frac{\sigma^2}{n}$

$$\text{Des.V}[\bar{x}] = \sqrt{V[\bar{x}]} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{30}{\sqrt{36}} = 5$$

c) $f(x) \sim$ Desconocida

c) $f(x) \sim$ Desconocida

tamaño de la muestra $\triangleq n = 36$

$\therefore$ como $n > 36$ por el teorema de limite central

$$\bar{x} \dot{\sim} \text{Normal}\left(\mu_{\bar{x}} = 500,\ \sigma^2_{\bar{x}} = (5)^2\right)$$

6]

- $f(x) \sim d?$ • $\mu = 700$ • $\sigma = 120$ • $n = 150$

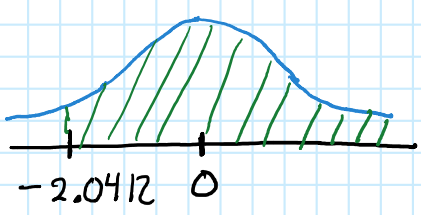
$\rightarrow P(\bar{x} > 680) =$

- $\mu_{\bar{x}} = E[\bar{x}] = 700$
- $\sigma^2_{\bar{x}} = V[\bar{x}] = \frac{120^2}{150} = \frac{\sigma^2}{n}$; $\sigma_{\bar{x}} = \frac{120}{\sqrt{150}} = \frac{\sigma}{\sqrt{n}}$

// Estandarizando

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_x} = \frac{\bar{x} - \mu_x}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{680 - 700}{\left(\frac{120}{\sqrt{150}}\right)} = -2.0412$$

$\therefore P(z > -2.0412) = .9793$



24]

• $\sigma = 25$ • Muestras de tamaño $= 16, 25, 50$ y 100

• $ES[\bar{x}] = \frac{\sigma}{\sqrt{n}}$ // Mientras mayor muestra $\triangleq$ menos error

• if n

16	25	50	100
$\frac{25}{\sqrt{16}}$	$\frac{25}{\sqrt{25}}$	$\frac{25}{\sqrt{50}}$	$\frac{25}{\sqrt{100}}$
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
6.25	5	3.5355	2.5

26]

26. Se aceptará un cargamento de barras de acero para el laboratorio de manufactura si la resistencia media a la ruptura de una muestra aleatoria de 10 barras es mayor que **235** libras por pulgada cuadrada. Por experiencia se sabe, la resistencia media a la ruptura de tales barras ha tenido promedio y desviación estándar iguales a **240** y 40, respectivamente.

a) Supóngase que la resistencia a la ruptura está distribuida normalmente, ¿cuál es la probabilidad de que una barra seleccionada aleatoriamente tenga una resistencia en el intervalo de 245 a 255?

b) ¿Cuál es la probabilidad de que el cargamento que llega al laboratorio sea aceptado?

DATOS

~~•~~ $\mu_{\bar{x}} > 235$ [libras^2]

• Muestra $= 10$ barras

• $\mu_x = 240$ • $\sigma_x = 40$

• $\mu_{\bar{x}} = E[\bar{x}] = 240$

• $ES[\bar{x}] = \frac{40}{\sqrt{10}}$

$\rightarrow V[\bar{x}] = \frac{\sigma^2}{n} = 160$

a) $P(245 < \bar{x} < 255) =$

$= P\left(\frac{245-240}{\frac{40}{1}} < z < \frac{255-240}{\frac{40}{1}}\right) = P(.125 < z < .375) =$

$= P(z < .0375) - P(z < .125) =$

$= .6461 - .5497 = .0964$ //

b) $f(x) \sim N(240, 40^2)$

$\therefore P(X > 235) = 1 - P(X < 235) =$

$= 1 - P\left(Z < \frac{235 - 240}{\sqrt{160}}\right) =$

$= 1 - P(Z < -.395) = .6535$ //

Estimadores - Ejercicios

Thursday, March 7, 2019 9:39 AM

- Para encontrar estimadores para algún parametro y existen métodos:
  - ↳ Media, Varianza, Proporción
  - → Método de los Momentos
  - → Método de Máxima verosimilitud

32] **32.** Un fabricante de bombillas afirma que sus productos durarán un promedio de 500 horas de trabajo. Para conservar este promedio esta persona verifica 25 focos cada mes. Si el valor y calculado cae entre –t 0.05 y t 0.05, él se encuentra satisfecho con esta afirmación. ¿Qué conclusión deberá él sacar de una muestra de 25 focos cuya duración fue?

520 521 511 513 510 513 522 500 521 495 496 488 500
502 512 510 510 475 505 521 506 503 487 493 500

Muestra $\triangleq n = 25$

- Suponiendo que la distribución de la población sea normal y que se realizó una muestra aleatoria simple

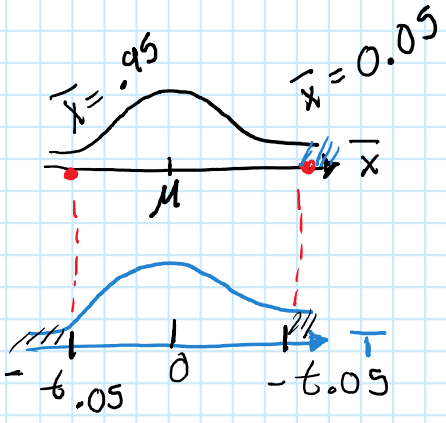
NOTA

$Y \sim \text{Normal}(\mu_Y, \sigma_Y^2)$

$Z = \dfrac{Y - \mu_Y}{\sigma_Y}$ $\quad T = \dfrac{Y - \mu_Y}{S_Y}$

$T \sim$ T - student

// σ^2 es desconocida



$= t = .95$

$T = \dfrac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$

$t_{0.05,24} = 1.711$

$t_{0.95,24} = -1.711$

Grados libertad $n-1$

→ De la muestra se tiene:

Conclusión:

→ De la muestra se tiene:

$$\left.\begin{array}{l}\bar{X} = 505.36 \\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\end{array}\right\} = T_{muestral}$$

$s^2 = 145.57$; $T_{muestral} = 2.22$ //

Conclusión:

Están dando mayor duración de la esperada en sus productos.

▶ 28

$P(|\hat{p} - \bar{p}| \leq 0.19) = ?$

$\sigma_p^2 = \frac{\hat{p}q}{n}$

↳ $\sigma_p^2 = \hat{p}q$

- Estadística descriptiva
- Conceptos de Inferencia Estadística
- Estimación Puntual
  - Métodos para encontrar estimador
  - Criterios para evaluar estimadores

Método de Máxima Verosimilitud

ALUMNO : Murrieta Villegas Alfonso

Monday, March 11, 2019 10:22 AM

Verosimilitud se define como la apariencia de verdad y es creible por no ofrecer ningún caracter de falsedad.
Si la variable aleatoria $\bar{x}$ tiene una distribución de probabilidad con algún parámetro θ, se asume que la forma de la función de densidad se conoce pero se desconoce el valor de θ; si se saca una muestra aleatoria de n observaciones independientes $(x_1, x_2, x_3 \ldots, x_n)$

$$v(\theta) = L(x_1, x_2, \ldots, x_n | \theta)$$

Define la función verosimilitud o densidad de probabilidad de los datos de la muestra dados el parametro θ.
Puesto que $v(\theta)$ es una función de θ ya que se conocen los valores de la muestra, para cada posible valor de θ la verosimilitud
El principio dice que cuando estamos puestos con varios valores de parametros, alguno de los cuales es verdadero de la población.

Ejemplo: Si el movimiento de precios de algún producto se comporta como un proceso de Bernoulli e importa si el precio diario o no sube, π es probabilidad de que los precios suban, se considera que π permanece constante día a día, que los cambios

se considera que π permanece constante día a día, que los cambios de los precios es independiente

3 posibles valores del parametro π, 0.4 y .6 y el resultado de la muestra fue en 9 de los 15 días. Entonces considerando la distancia binomial

$$\pi = .4 \rightarrow P(R=9 / n=15, \pi .04) = \binom{15}{9}(.4)^9(.6)^6 = .061$$

$$\pi = .6 \rightarrow P(R=9 / n=15, \pi=.6) = \binom{15}{9}(.6)^9(.4)^6 = .207$$

π más probable

$$v(\pi) = [(x_1, x_2, x_3 \ldots, x_n)\pi] = \binom{n}{r}(\pi)^r(1-\pi)^{n-r}$$

Esta es una dist. sobre las posibles muestras conocidas los valores de los parámetros de interés

El uso se basa en que la evidencia completa de la muestra este contenido en la función y cuando se saca la muestra el investigador no está interesado en la probabilidad de otras muestras

Cuando se encuentre la función, que en el ejemplo el siguiente paso es encontrar los valores del parametro

$$v(\pi) = \binom{n}{r}(\pi)^r(1-\pi)^{n-r}$$

$$\log(v(\pi)) = \log\left[\binom{n}{r}(\pi)^r(1-\pi)^{n-r}\right]$$

$$\frac{d \log[v(\pi)]}{d\pi} = \log\binom{n}{r} + r\log(\pi) + (n-r)(\log \pi)$$

$$\frac{r}{n} - \frac{n-r}{1-\pi} = \frac{r(1-\pi) - \pi(n-r)}{} = 0$$

$$\frac{r}{n} - \frac{n-r}{1-\pi} = \frac{r(1-\pi) - \pi(n-r)}{\pi(1-\pi)} = 0$$

$$\therefore\ r(1-\pi) - \pi(n-r) = r - r\pi - n\pi + r\pi = r - n\pi = 0$$

$$\pi = \frac{r}{n}$$

Referencia

- Estimación Estadística de parámetros.
  dcb.fi-c.unam.mx/profesor/irene/BEP/capobfc/cap17-bfc2001.pdf

---

$X = \{X_1, X_2, \ldots, X_n\}$

una muestra de una población con distribución $f(x,\theta)$

entonces $f_{X_i}(x_i, \theta) = f(x, \theta)$

Distribución de Probabilidad de cada observación de la muestra = Distribución Poblacional

La función de verosimilitud de la muestra es la probabilidad conjunta de todas las observaciones de la muestra

$\therefore$ Consiste en encontrar aquellos valores de los parámetros que maximizan la función de verosimilitud

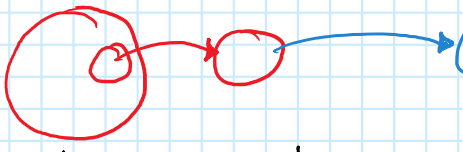
$$L(x) = \prod_{i=1}^{n} f(x_i)$$

Review

Tuesday, March 12, 2019 9:16 AM

Inferencia Estadística

- ¿Cómo es la población? // $f(x, \theta)$
- ¿Cómo son los valores de los parametros poblacionales?



Población ; Muestra

H Estadístico $\Rightarrow \bar{x} = \frac{1}{n} \sum x$

$\hat{\theta}$ Estimador $\Rightarrow E\{\bar{x}\} = \mu$

$\therefore \bar{x} = \hat{\mu}$

Para poder hacer estimaciones de algún parametro se requiere:

1] Definir cuál es el parámetro de interés

2] Tener un estimador para un parámetro $\hat{\theta}$

3] Saber cuál es la distribución de muestreo de $\hat{\theta}$ y sus características

Distribución de muestreo | Es la distribución de probabilidad de un estadístico.

Técnicas para encontrar estimadores puntuales

a) Método de los momentos

b) Método de máxima verosimilitud

Criterios para evaluar estimadores

1) Sesgo → Estimador ; Diferencia del estimador ,

1] Sesgo → Estimador insesgado | Diferencia del estimador respecto al valor esperado

2] Varianza

3] Eficiente → Es aquel que tenga la menor varianza entre todos los estimadores existentes

4] Consistente

5] Suficiente

Tanto para el método de los momentos como para el de verosimilitud, el estimador se encuentra para σ^2 es: $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ // Insesgado

ya que $E\{S_n^2\} = \sigma^2$

Entonces la opción para estimar sigma sería sustituir μ por $\bar{x}$

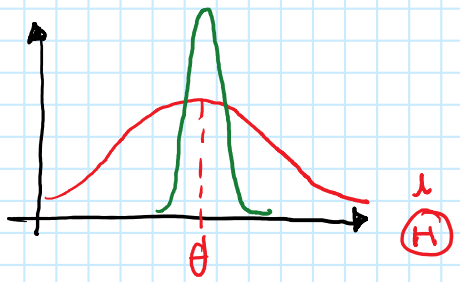
$$S_{2n}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Pero resulta que | Es sesgado

$E\{S_{2n}^2\} \neq \sigma^2$ | ∴ Se buscará corregir el sesgo

Varianza Muestral | $\sigma_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ | ¿Cuál es el mejor estimador de la varianza poblacional?

// Porque tiene corregido el sesgo



- es un mejor estimador debido a tener menor varianza

Error Cuadrático Medio

- $ECM(\hat{\Theta}) = E\{(\hat{\Theta} - \theta)^2\}$

//Desarrollando

- $ECM(\hat{\Theta}) = Var\{\hat{\theta}\} + (sesgo\{\hat{\theta}\})^2$

NOTA | Los estadísticos no deben tener datos desconocidos

Problema 18
- (Archivo 7)

Sea $X_1, X_2, \ldots X_7$ una muestra aleatoria de una población que tiene media μ y varianza σ^2

- $\hat{\Theta}_1 = \dfrac{X_1 + X_2 + \ldots X_7}{7}$

$$E\{\hat{\theta}_1\} = E\left\{\frac{1}{7}\sum_{i=1}^{7} X_i\right\} = \frac{1}{7}\sum_{i=1}^{7} E\{X_i\} = \frac{1}{\cancel{7}}\cancel{7}\mu = \mu$$

$\therefore$ θ_1 es insesgado //

- $\hat{\Theta}_2 = \dfrac{2X_1 - X_6 + X_7}{2}$

$$E\{\theta_2\} = E\left\{\frac{2X_1 - X_6 + X_7}{}\right\} = \frac{1}{2}\left[2E\{X\} - E\{X_6\} + E\{X_7\}\right]$$

$$E\{\theta_2\} = E\left\{\frac{2x_1 - x_6 + x_7}{2}\right\} = \frac{1}{2}\left[2E\{x\} - E\{x_6\} + E\{x_7\}\right]$$

$$= \frac{1}{2}\, 2\mu = \mu$$

$\therefore \hat{\theta}_2$ es insesgado //

- $V\{\hat{\theta}\} = Var\left\{\frac{1}{7}\sum_{i=1}^{7} x_i\right\} = \frac{1}{7^2}\sum_{i=1}^{7} Var\{x\} =$

$$= \frac{1}{7^2}\, 7\,\sigma^2 = \frac{\sigma^2}{7}$$

- $V\{\hat{\theta}_2\} = \frac{1}{2^2}\left[4\sigma^2 + \sigma^2 + \sigma^2\right] = \frac{6}{4}\sigma^2 = \frac{3}{2}\sigma^2$

// Eficiencia relativa

$$\eta_{\hat{\theta}_1/\hat{\theta}_2} = \frac{Var\{\hat{\theta}_1\}}{Var\{\hat{\theta}_2\}} = \frac{\frac{\sigma^2}{7}}{\frac{3\sigma^2}{2}} = \frac{2}{21} = 0.095$$

Proporción de eficiencia respecto uno de otro

Estimadores Puntuales

- Para encontrar estimadores: { método de momentos / método de máxima verosimilitud
- Para evaluar los estimadores: { - Insesgado / - Eficiencia Mod

- Para evaluar los estimadores:
  - Insesgado
  - Error Cuadrático Medio
  - Óptimo → De varianza menor
  - Eficiente
  - Consistente
  - Suficiente

▶ Neymann-Pearson

Máxima Verosimilitud | $L(\underline{x}, \theta) = h(t; \theta)\, g(\underline{x})$

Ejercicios

Thursday, March 14, 2019 9:28 AM

6.- Los costos de afinación en una estación de servicio grande son de 30 U.M. para un coche de cuatro cilindros, de 36 para uno de seis cilindros y de 42 para uno de ocho cilindros; de los registros hallados se sabe que el 10% de las afinaciones se hacen para coches de cuatro cilindros, el 40% para coches de seis cilindros y el 50% para coches de ocho cilindros. Se seleccionan al azar dos coches para afinación.

a) Construir la tabla de distribución de probabilidad para el promedio del costo del servicio.

b) Obtener $\mu_{\bar{X}}$.

Respuesta: b) 38.4

- $X \triangleq$ Costo de afinación

X	$P(x)$
30	0.1
36	0.4
42	0.5

$$\mu = \sum x_i P(x_i) = 38.4$$

- $\sigma^2 = \mu_2 - \mu^2 = \left[\sum x_i^2 P(x_i)\right] - \mu^2 = 15.84$

- Tabla en Excel → Checar

$$P(\bar{x} = 33) = P(x_1 = 30 \cap x_2 = 36) \cup P(x_1 = 36 \cap x_2 = 30)$$

c) Encontrar la Varianza de la Media Muestral

$$\sigma^2_{\bar{x}} = Var\{\bar{x}\} = \left[\sum \bar{x}^2 \cdot P(\bar{x})\right] - \mu^2_{\bar{x}}$$

Varianza poblacional entre el número de muestras

7.- Un juguete está formado por tres partes, cuyos respectivos pesos (en gramos) son las variables aleatorias X_1, X_2 y $X_3 X_3$. Asumiendo que los pesos son independientes y con las distribuciones normales con medias y variancias:

$$X_1 \sim N(150, 36), \quad X_2 \sim N(100, 25), \quad X_3 \sim N(250, 60)$$

7.- Un juguete está formado por tres partes, cuyos respectivos pesos (en gramos) son las variables aleatorias X_1, X_2 y $X_3 X_3$. Asumiendo que los pesos son independientes y con las distribuciones normales con medias y variancias:

$$X_1 \sim N(150,\ 36),\quad X_2 \sim N(100,\ 25),\quad X_3 \sim N(250,\ 60)$$

a) Obtener la distribución y los parámetros del peso total T.

b) Determinar la proporción de juguetes con peso mayor que 520 gramos.

c) Considerando que se guardan 24 juguetes en una caja que pesa 200 gramos, calcular el valor esperado y la variancia del peso total de la caja con los juguetes.

d) ¿Cuál es la probabilidad de que el peso de la caja con los juguetes sea mayor que 12300 gramos?

Respuestas: b) 0.0344, d) 0.0314

• Construir la Distribución de Muestreo de $\bar{X}$ cuando $n=2$

a)

$X_1, X_2, X_3 \sim$ Normal

$\therefore T = X_1 + X_2 + X_3 \sim$ Normal

$\rightarrow \mu_T = E\{X_1\} + E\{X_2\} + E\{X_3\} =$

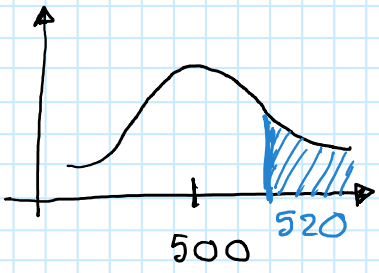
$= 150 + 100 + 250 = 500$ gramos

$\rightarrow \sigma_T^2 = V\{X_1\} + V\{X_2\} + V\{X_3\}$

$= 36 + 25 + 60 = 121$

$T \sim \text{Normal}(500, 121)$

b)



$T \triangleq$ Peso total del juguete

$P(T > 520) = 1 - P(T < 520) = .03451$

$= 1 -$ dist. n.n. $n(520, 500, \sigma = \sqrt{121})$

$P(Y > 520) = 1 - P(Y < 520) = .03451$ //

$= 1 - \text{distr.norm.n}(520, 500, \text{raiz}(121), 1)$

Acumulativa

Nota: $\sqrt{121}$ = Desviación Estandar

// Estandarización

$$1 - P\left(Z < \frac{520-500}{\sqrt{121}}\right)$$

c) $W \triangleq$ Peso de los 24 juguetes y el de la caja

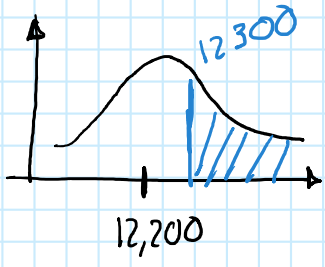
$W = \sum_{i=1}^{24} T_i + C$ $\qquad C \triangleq$ Peso de la caja

$\therefore W \sim \text{Normal}(\overset{12,200}{\cancel{24 * 500 + 200}}, \overset{2904}{\cancel{24 * 121}})$

$\sim \text{Normal}(\underbrace{12,200}_{\mu_W}, \underbrace{2904}_{\sigma_W^2})$

$\mu_W = 12,200$ // $\qquad \sigma_W^2 = 2904$ //

d)



$P(W > 12,300) = 1 - P(W < 12,300) =$

$= .03175$ //

28] **28.** El Coordinador de Ciencias Aplicadas, desea recabar información sobre la proporción de alumnos a los que no les agrada una nueva política respecto a la aceptación de alumnos irregulares. ¿Cuántos alumnos debe incluir en una muestra si desea que la fracción de la muestra se desvíe a lo más en 0.15 de la verdadera fracción, con una probabilidad de 0.98?

Error : $P(|\hat{p} - p| < .15) = .98$;

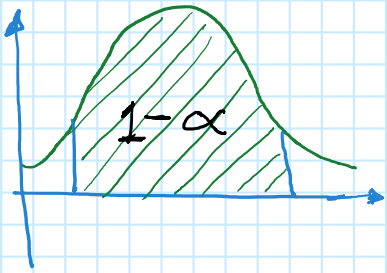
Estimador - Real

$$P(-.15 < \hat{p} - p < .15) = .98$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \triangleq \text{Tiene distribución normal}$$

$$E = |\hat{p} - p| = .15$$

$$\therefore P\left(\frac{-E}{\sqrt{\frac{pq}{n}}} < \hat{p} - p < \frac{E}{\sqrt{\frac{pq}{n}}}\right) = .98;$$



$$\text{if} \quad z_{\frac{\alpha}{2}} = \frac{E}{\sqrt{\frac{pq}{n}}} \; ; \; n = \frac{\left(z_{\frac{\alpha}{2}}\right)^2 pq}{(E)^2}$$

$$\alpha = 1 - .98 = .2$$

$$\frac{\alpha}{2} = .1$$

Distribución normal;

if $p = .5$ para mayor valor

$$n = \frac{(2.326^2)(.5)(.5)}{(.15)^2} \cong 60.11$$

9.- El peso medio de equipaje registrado por un pasajero de clase turista seleccionada al azar, para cierto vuelo de una aerolínea, es de 40 [kg] y la desviación estándar es 10 [kg]. La media y la desviación estándar del peso del equipaje para un pasajero de clase ejecutiva son 25 y 6 [kg], respectivamente.

a) Si hay 12 pasajeros de clase ejecutiva y 50 pasajeros clase turista en un vuelo en particular, determinar los valores de la media y la desviación estándar del peso total del equipaje.

b) Si los pesos individuales de equipaje son variables aleatorias independientes y normalmente distribuidos, determinar la probabilidad de que el peso total del equipaje sea a lo sumo 2500 [kg].

Respuesta: b) 0.9966

- $X_1 \triangleq$ Peso de equipaje por pasajero (Turista)
- $X_2 \triangleq$ Peso de equipaje por pasajero (Ejecutiva)

$\mu = 40$ [kg] $\sigma = 10$ | $\mu = 25$ [kg] $\sigma = 6$

$\mu_1 = 40\,[kg] \quad \sigma_1 = 10 \;\Big|\; \mu_2 = 25\,[kg] \quad \sigma_2 = 6$

- $Y \triangleq$ Peso Total

$$Y = \sum_{i=1}^{50} X_1 + \sum_{i=12}^{12} X_2$$

$$\rightarrow E[Y] = \sum_{i=1}^{50} E[X_1] + \sum_{i=1}^{12} E[X_2] =$$

$$= 50(40) + 12(25) = 2300$$

$$\rightarrow V[Y] = \sum_{i=1}^{50} V\{X_1\} + \sum_{i=1}^{12} V\{X_2\}$$ //Varianza no desviación estandar

$$= 50(10)^2 + 12(6)^2 =$$

$$= 5432$$

Desviación Estandar $\Big|$ $\sigma_Y = \sqrt{V\{Y\}} = \sqrt{5432} = 73.7$

Parámetros de Y $\Big|$ $\mu_Y = 2300$, $\sigma_Y = 73.7$

b) $P(Y \leq 2500) =$

$$P \sim Normal(\mu_Y = 2300,\ \sigma_Y^2 = 5432)$$

$$= Norm.Dis(2500, 2300, 73.7, 1)$$

$$= .9966$$

$= .9966$

$$= P\left(Z \leq \frac{2500-2300}{\sqrt{5432}}\right) = P\left(Z \leq 2.71\right) =$$

$$= .9966$$

17]

17.- Supóngase que cierto tipo de fertilizantes tiene una producción esperada por acre de μ_1 con variancia σ^2, en tanto que la producción esperada para un segundo tipo de fertilizante es μ_2 con la misma variancia σ^2. Denótense por S_1^2 y S_2^2 las variancias muestrales insesgadas de producciones basadas en tamaños n_1 y n_2, respectivamente, de los dos fertilizantes. Demostrar que el estimador combinado

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad \text{es insesgado para } \sigma^2.$$

Suponemos que las varianzas poblacionales son iguales

$$\sigma_1^2 = \sigma_2^2 = \hat{\sigma}$$

$$\begin{cases} S_1^2 = \sigma_1^2 = \hat{\sigma} \\ S_1^2 = \sigma_2^2 = \hat{\sigma} \end{cases}$$

$$E\{\hat{\sigma}^2\} = E\left\{\frac{(n-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right\} =$$

$$= \frac{1}{n_1+n_2-2} E\left\{(n-1)S_1^2 + (n_2-1)S_2^2\right\}$$

$\hat{\sigma}_2 \doteq$ Es un estimador insesgado

2_Examen

Thursday, March 14, 2019 1:12 PM

ALUMNO: MURRIETA VILLEGAS ALFONSO			EXAMEN SORPRESA: 2

8.- Si el error estándar de la media para la distribución muestral de las muestras aleatorias de tamaño 36 de una población grande o infinita es 2, ¿qué tan grande debe ser el tamaño de la muestra si el error estándar debe reducirse a 1.2?
Respuesta: 100

Datos

- $n = 36$
- $P = \infty$
- Error Estandar $\triangleq ES_{\bar{x}} = 2$

Expresión

$$ES = \frac{\sigma}{\sqrt{n}}$$

// Obteniendo σ

$$\frac{\sigma}{\sqrt{n}} = 2 \; ; \; \sigma = (2)(\sqrt{36}) = 12 \; ; \; \sigma = 12$$

// Reduciendo ES a 1.2

$$1.2 = \frac{\sigma}{\sqrt{n}} \; ; \; \sqrt{n} = \frac{\sigma}{1.2} \; ; \; n = \left(\frac{12}{1.2}\right)^2$$

$\therefore$ El tamaño de la muestra $n = 100$